2014 American Control Conference (ACC)
June 4-6, 2014. Portland, Oregon, USA

Distributed Estimation using Bayesian Consensus Filtering

Saptarshi Bandyopadhyay *Student Member, IEEE*, Soon-Jo Chung *Senior Member, IEEE*

***Abstract*—We present the Bayesian consensus filter (BCF) for tracking a moving target using a networked group of sensing agents and achieving consensus on the best estimate of the probability distributions of the target's states. Our BCF framework can incorporate nonlinear target dynamic models, heterogeneous nonlinear measurement models, non-Gaussian uncertainties, and higher-order moments of the locally estimated posterior probability distribution of the target's states obtained using Bayesian filters. If the agents combine their estimated posterior probability distributions using a logarithmic opinion pool, then the sum of Kullback–Leibler divergences between the consensual probability distribution and the local posterior probability distributions is minimized. Rigorous stability and convergence results for the proposed BCF algorithm with single or multiple consensus loops are presented. Communication of probability distributions and computational methods for implementing the BCF algorithm are discussed along with a numerical example.**

I. Introduction

Distributed estimation, where a group of networked agents collectively estimate the target's states, can be used for environment and pollution monitoring, tracking dust or volcanic ash clouds, tracking orbital debris or asteroids in space, etc [1]– [5]. The term *distributed estimation* refers to finding the best estimate of the target's states using the sensor network while the term *consensus* means reaching an agreement across the network [6]– [10].

Many existing algorithms for distributed estimation [1]– [5], [11]– [15] aim to reach an agreement across the network on the estimated mean (first moment of the estimated probability distribution) of the target dynamics, but cannot incorporate nonlinear target dynamics, heterogeneous nonlinear measurement models, non-Gaussian uncertainties, or higher-order moments of the locally estimated posterior probability distribution of the target's states. It is difficult to recursively combine local mean and covariance estimates using a linear consensus algorithm because the dimension of the vector transmitted by each agent increases linearly with time due to correlated process noise [16] and the covariance update equation is usually approximated by a consensus gain [17].

Multi-agent tracking or sensing networks are deployed in a distributed fashion when the target dynamics have complex temporal and spatial variations. Hence, it is necessary to preserve the complete information captured in the locally estimated posterior probability distribution of the target's states while achieving consensus across the network. The main aim of this paper is to extend the scope of distributed estimation algorithms to track targets with general nonlinear dynamic models with stochastic uncertainties, thereby addressing the aforementioned shortcomings.

Bayesian filters [18], [19] recursively calculate the probability density/mass function of the beliefs and update them based on new measurements. The main advantage of Bayesian filters over Kalman filter–based methods for estimation of nonlinear target dynamic models is that no approximation is needed during the filtering process. Advances in computational capability have facilitated the implementation of Bayesian filters for robotic localization and mapping [20] as well as planning and control [21]. Practical implementation of these algorithms, in their most general form, is achieved using particle filtering [22] and Bayesian programming [23]. This paper focuses on developing a consensus framework for distributed Bayesian filters.

The *statistics* literature deals with the problem of reaching a consensus among individuals in a complete graph, where each individual's opinion is represented as a probability distribution [24], [25]; and under select conditions, it is shown that consensus is achieved within the group [26]. Exchange of beliefs in decentralized systems, under communication constraints, is considered in [27], [28]. Algorithms for combining probability distributions within the exponential family, i.e., the limited class of unimodal distributions that can be expressed as an exponential function, are studied in [29], [30]. If the target's states are discrete random variables, then the local estimates can be combined using a tree-search algorithm [31] or a linear consensus algorithm [32], [33]. In contrast, this paper focuses on developing generalized Bayesian consensus algorithms with rigorous convergence analysis for achieving consensus across the network without any assumption on the shape of local prior or posterior probability distributions. The proposed distributed estimation using Bayesian consensus filtering aims to reach an agreement across the network on the best estimate, in information theoretic sense, of the probability distribution of the target's states.

In this paper, we assume that agents generate their local estimate of the posterior probability distribution of the target's states using Bayesian filters with/without measurement exchange with neighbors. Then, we develop algorithms for combining these local estimates, using the logarithmic opinion pool (LogOP), to generate the consensual estimate of the probability distribution of the target's states across the network.

The authors are with the Department of Aerospace Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. (email: bandyop2@illinois.edu; sjchung@illinois.edu)

This research was supported by AFOSR grant FA95501210193. An extended version of this paper has been submitted to IEEE Transactions on Automatic Control. (http://arxiv.org/abs/1403.3117)

Finally, we introduce the Bayesian consensus filter (BCF), where the local prior estimates of the target's states are first updated and the local posterior probability distributions are recursively combined during the consensus stage, so that the agents can estimate the consensual probability distribution of the target's states while simultaneously maintaining consensus across the network. The pseudo-code for the algorithm is given in **Algorithm 1**. The novel features of the BCF algorithm are:

- The algorithm can be used to track targets with general nonlinear time-varying target dynamic models.
- The algorithm can be used by a SC balanced network of heterogeneous agents, with general nonlinear time-varying measurement models.
- The algorithm achieves global exponential convergence across the network to the consensual probability distribution of the target's states.
- The consensual probability distribution is the best estimate, in the information theoretic sense because it minimizes the sum of KL divergences with the locally estimated posterior probability distributions. If a central agent receives all the local posterior probability distributions and is tasked to find the best estimate in the information theoretic sense, then it would also yield the same consensual probability distribution. Hence, we claim to have achieved distributed estimation using the BCF algorithm.

The Hierarchical BCF algorithms is used when some of the agents do not observe the target. We apply the Hierarchical BCF algorithm to the problem of tracking orbital debris in space using the space surveillance network on Earth.

A. Notation

The *time index* is denoted by a right subscript while the *agent index* is denoted by a lower-case right superscript. Let $\boldsymbol{x}_k$ represent the true target's states at the k^{th} time instant. Let $\boldsymbol{z}_k^j$ represents the measurement taken by the j^{th} agent at the k^{th} time instant. Let $\mathcal{F}_k^j$ represents the estimated probability density function (pdf) of the target's states over the state space $\mathcal{X}$. The symbol $p(\cdot)$ also refers to pdf over the state space $\mathcal{X}$.

The graph $\mathcal{G}_k$ represents the directed time-varying communication network topology at the k^{th} time instant, where all the agents form the set of vertices $\mathcal{V}$ and $\mathcal{E}_k$ is the set of directed edges. Let $\mathcal{N}_k^j$ denote the neighbors of the j^{th} agent at the k^{th} time instant from which it receives information and $\mathcal{J}_k^j := \mathcal{N}_k^j \cup \{j\}$ denote the set of inclusive neighbors.

Let $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{R}^{m\times n}$ be the sets of natural numbers (positive integers), real numbers, and m by n matrices. Let λ and σ represent the eigenvalue and the singular value of a square matrix. Let $\mathbf{1} = [1, 1, \ldots, 1]^T$, $\mathbf{I}$, and $\mathbf{0}$ be the ones vector, the identity matrix, and the zero matrix of appropriate sizes. The symbols $|\cdot|$, $\lceil\cdot\rceil$, $\ln(\cdot)$, and $\log_c(\cdot)$ represent the absolute value, ceiling function, the natural logarithm and the logarithm to the base c. Finally, $\|\cdot\|_{\ell_p}$ represents the ℓ_p vector norm. The $\mathcal{L}_p$ function denotes the set of all functions $f(\boldsymbol{x}) : \mathbb{R}^{n_x} \to \mathbb{R}$ with the bounded integral $\left(\int_{\mathcal{X}} |f(\boldsymbol{x})|^p d\mu(\boldsymbol{x})\right)^{1/p}$, where μ is a measure on $\mathscr{X}$.

II. Preliminaries

In this section, we first state four assumptions used throughout this paper and then introduce the problem statement of BCF. Next, we discuss an extension of the Bayesian filter to sensor fusion over a network.

Assumption 1. In this paper, all the algorithms are presented in discrete time. □

Assumption 2. The state space ($\mathcal{X} \subset \mathbb{R}^{n_x}$) is closed and bounded, hence $\mathcal{X}$ is compact. □

Assumption 3. All continuous probability distributions are upper-bounded by some large value $\mathcal{M} \in \mathbb{R}$. □

Assumption 4. The inter-agent communication time scale is much faster than the tracking/estimation time scale. □

These assumptions have been introduced to make the algorithms computationally tractable and to take advantage of the results dealing with bounded functions and compact support. Note that discrete, continuous or mixed probability distribution can be handled in a unified manner using measures. Hence, every probability distribution in this paper is expressed as a probability density function (pdf) over $\mathcal{X}$.

Let $\mathcal{X} \subset \mathbb{R}^{n_x}$ be the n_x-dimensional state space of the target. The dynamics of the target in discrete time $\{\boldsymbol{x}_k, k \in \mathbb{N}, \boldsymbol{x}_k \in \mathcal{X}\}$ is given by:

$$\boldsymbol{x}_k = \boldsymbol{f}_k(\boldsymbol{x}_{k-1}, \boldsymbol{v}_{k-1}) , \tag{1}$$

where $\boldsymbol{f}_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_x}$ is a possibly nonlinear time-varying function of the state $\boldsymbol{x}_{k-1}$ and an independent and identically distributed (i.i.d.) process noise $\boldsymbol{v}_{k-1}$, where n_v is the dimension of the process noise vector. Let m heterogeneous agents simultaneously track this target and estimate the pdf of the target's states (where m does not change with time). The measurement model of the j^{th} agent is given by:

$$\boldsymbol{z}_k^j = \boldsymbol{h}_k^j(\boldsymbol{x}_k, \boldsymbol{w}_k^j), \quad \forall j \in \{1, \ldots, m\}, \tag{2}$$

where $\boldsymbol{h}_k^j : \mathbb{R}^{n_x} \times \mathbb{R}^{n_{wj}} \to \mathbb{R}^{n_{zj}}$ is a possibly nonlinear time-varying function of the state $\boldsymbol{x}_k$ and an i.i.d. measurement noise $\boldsymbol{w}_k^j$, where n_{zj}, n_{wj} are dimensions of the measurement and measurement noise vectors respectively. Note that the measurement model of agents is quite general since it accommodates heterogeneous sensors, partial state observation, varied sensors for sensor fusion in frequency domain, etc.

The objective of the BCF is to estimate the target's states and maintain consensus across the network. This objective is achieved in two steps: (i) each agent locally estimates the pdf of the target's states using a Bayesian filter, and (ii) each agent's local estimate converges to a global estimate during the consensus stage.

A. Bayesian Filter with Measurement Exchange

The objective of Bayesian filtering with/without measurement exchange is to estimate the posterior pdf of the target's states at the k^{th} time instant, which is denoted by $\mathcal{F}_k^j, \forall j \in \{1, \ldots, m\}$, using the estimated prior pdf of the target's states $\mathcal{F}_{k-1}^j$ from the $(k-1)^{\text{th}}$ time instant and the new measurement array obtained at the k^{th} time instant. Exchange of measurements is optional since heterogeneous agents, with different priors, fields of view, resolutions, tolerances, etc., may not be able to combine measurements from other agents. For example, if a satellite in space and a low flying quadrotor are observing the same target, then they cannot exchange measurements due to their different fields of view.

If an agent can combine measurements from another neighboring agent during its update stage, then we call them *measurement neighbors*. In this section, we extend the Bayesian filter by assuming that each agent transmits its measurements to other agents in the network, and receives the measurements from its measurement neighbors. Let $\boldsymbol{z}_k^{\mathcal{S}_k^j} := \{\boldsymbol{z}_k^\ell, \forall \ell \in \mathcal{S}_k^j\}$ denote the array of measurements taken at the k^{th} time instant by the measurement neighbors of the j^{th} agent, where $\mathcal{S}_k^j \subseteq \mathcal{J}_k^j$ denotes the set of measurement neighbors among the inclusive neighbors of the j^{th} agent.

It is assumed that each agent either knows the prior $\mathcal{F}_0^j = p(\boldsymbol{x}_0|\boldsymbol{z}_0^j)$ or $\mathcal{F}_0^j$ is assumed to be uniformly distributed over $\mathcal{X}$. The prediction stage involves using the target dynamics model (1) to obtain the estimated pdf of the target's states at the k^{th} time instant via the Chapman–Kolmogorov equation:

$$p_k^j(\boldsymbol{x}_k) = \int_{\mathcal{X}} p_k^j(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})\, p_{k-1}^j(\boldsymbol{x}_{k-1})\, d\mu(\boldsymbol{x}) \quad (3)$$

The probabilistic model of the state evolution $p_k^j(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ is defined by the target dynamics model (1) and the known statistics of the i.i.d. process noise $\boldsymbol{v}_{k-1}$. The new measurement array ($\boldsymbol{z}_k^{\mathcal{S}_k^j}$) is used to compute the posterior pdf of the target's states ($\mathcal{F}_k^j = p_k^j(\boldsymbol{x}_k|\boldsymbol{z}_k^{\mathcal{S}_k^j})$) during the update stage using Bayes' rule (4):

$$p_k^j(\boldsymbol{x}_k|\boldsymbol{z}_k^{\mathcal{S}_k^j}) = \frac{\left(\prod_{\ell\in\mathcal{S}_k^j} p_k^\ell(\boldsymbol{z}_k^\ell|\boldsymbol{x}_k)\right) p_k^j(\boldsymbol{x}_k)}{\int_{\mathcal{X}}\left(\prod_{\ell\in\mathcal{S}_k^j} p_k^\ell(\boldsymbol{z}_k^\ell|\boldsymbol{x}_k)\right) p_k^j(\boldsymbol{x}_k)\, d\mu(\boldsymbol{x})}, \quad (4)$$

The likelihood function $p_k^\ell(\boldsymbol{z}_k^\ell|\boldsymbol{x}_k)$, $\forall \ell \in \mathcal{S}_k^j$ is defined by the measurement model (2), and the corresponding known statistics of the i.i.d. measurement noise $\boldsymbol{w}_k^\ell$.

Note that (4) is similar to the empirical equation for *Independent Likelihood Pool* given in [34] and a generalization of the *Distributed Sequential Bayesian Estimation Algorithm* given in [35]. The structure of (4) ensures that an arbitrary part of the prior distribution does not dominate the measurements.

III. Combining Probability Distributions

In this section, we present algorithms for achieving consensus in probability distributions across the network. The objective of the consensus stage in **Algorithm 1** is to guarantee pointwise convergence of each $\mathcal{F}_k^j$ to a consensual pdf $\mathcal{F}_k^\star$, which is independent of j. This is achieved by each agent recursively transmitting its estimated pdf of the target's states to other agents, receiving estimates of its neighboring agents, and updating its estimate of the target. Let $\mathcal{F}_{k,0}^j = \mathcal{F}_k^j$ represent the local estimated posterior pdf of the target's states, by the j^{th} agent at the start of the consensus stage, obtained using Bayesian filters with/without measurement exchange. During each of the n_{loop} iterations within the consensus stage in **Algorithm 1**, this estimate is updated as follows:

$$\mathcal{F}_{k,\nu}^j = \mathcal{T}\left(\cup_{\ell\in\mathcal{J}_k^j}\{\mathcal{F}_{k,\nu-1}^\ell\}\right), \forall j \in \{1,\ldots,m\}, \forall \nu \in \mathbb{N}, \quad (5)$$

where $\mathcal{T}(\cdot)$ is the linear or logarithmic opinion pool for combining the pdf estimates. Note that the problem of measurement neighbors does not arise here since all pdfs are expressed over the complete state space $\mathcal{X}$.

Let $\mathcal{F}_1, \ldots, \lim_{n\to\infty}\mathcal{F}_n$, $\mathcal{F}^\star$ be real-valued measurable functions on $\mathcal{X}$, $\mathscr{X}$ be the Borel σ-algebra of $\mathcal{X}$, and $\mathscr{A}$ be any set in $\mathscr{X}$. Let $\mu_{\mathcal{F}_n}$, $\mu_{\mathcal{F}^\star}$ denote the respective induced measures of $\mathcal{F}_n$, $\mathcal{F}^\star$ on $\mathscr{X}$.

Lemma 1. *(Pointwise convergence implies convergence in Total Variation) If $\mathcal{F}_n$ converges to $\mathcal{F}^\star$ pointwise, i.e., $\lim_{n\to\infty}\mathcal{F}_n = \mathcal{F}^\star$ pointwise; then the measure $\mu_{\mathcal{F}_n}$ converges in TV to the measure $\mu_{\mathcal{F}^\star}$, i.e., $\lim_{n\to\infty}\mu_{\mathcal{F}_n} \xrightarrow{T.V.} \mu_{\mathcal{F}^\star}$.*

Proof: The proof follows from Scheffé's theorem and dominated convergence theorem [36, pp. 84]. ■

The first method of combining the estimates is motivated by the linear consensus algorithms widely studied in the literature [6]– [10]. The pdfs are combined using the Linear Opinion Pool (LinOP) of probability measures [24], [25]:

$$\mathcal{F}_{k,\nu}^j = \sum_{\ell\in\mathcal{J}_k^j} a_{k,\nu-1}^{j\ell}\mathcal{F}_{k,\nu-1}^\ell, \ \forall j \in \{1,\ldots,m\}, \forall \nu \in \mathbb{N}, \quad (6)$$

where $\sum_{\ell\in\mathcal{J}_k^j} a_{k,\nu-1}^{j\ell} = 1$ and the updated pdf $\mathcal{F}_{k,\nu}^j$ after the ν^{th} consensus loop is a weighted average of the pdfs of the inclusive neighbors $\mathcal{F}_{k,\nu-1}^\ell, \forall \ell \in \mathcal{J}_k^j$ from the $(\nu-1)^{\text{th}}$ consensus loop, at the k^{th} time instant. The LinOP solution is typically multimodal and depends on the assumption that the same 0-1 scale is used by every agent, so no clear choice for jointly preferred estimate emerges from it [25].

A. Consensus using the Logarithmic Opinion Pool

Note that $\mathcal{F}_{k,\nu}^j = p_{k,\nu}^j(\boldsymbol{x}_k), \forall \boldsymbol{x}_k \in \mathcal{X}$ represents the pdf of the estimated target's states by the j^{th} agent during the ν^{th} consensus loop at the k^{th} time instant. The LogOP is given as [37]:

$$\mathcal{F}_{k,\nu}^j = p_{k,\nu}^j(\boldsymbol{x}_k) = \frac{\prod_{\ell\in\mathcal{J}_k^j}\left(p_{k,\nu-1}^\ell(\boldsymbol{x}_k)\right)^{a_{k,\nu-1}^{j\ell}}}{\int_{\mathcal{X}}\prod_{\ell\in\mathcal{J}_k^j}\left(p_{k,\nu-1}^\ell(\boldsymbol{x}_k)\right)^{a_{k,\nu-1}^{j\ell}} d\mu(\boldsymbol{x})},$$
$$\forall j \in \{1,\ldots,m\}, \forall \nu \in \mathbb{N}, \quad (7)$$

where $\sum_{\ell\in\mathcal{J}_{k,\nu-1}^j} a_{k,\nu-1}^{j\ell} = 1$ and the integral in the denominator of (7) is finite. Thus the updated pdf $\mathcal{F}_{k,\nu}^j$ after the ν^{th} consensus loop is the weighted geometric average of the pdfs of the inclusive neighbors $\mathcal{F}_{k,\nu-1}^\ell, \forall \ell \in \mathcal{J}_k^j$ from the $(\nu-1)^{\text{th}}$ consensus loop, at the k^{th} time instant. The LogOP solution is typically unimodal and less dispersed, indicating a consensual estimate jointly preferred by the network [25]. The LogOP solution is invariant under under rescaling of individual degrees of belief, hence it preserves an important credo of uni–Bayesian decision theory; i.e., the optimal decision should not depend upon the choice of scale for the utility function or prior probability distribution. The most compelling reason for using LogOP is that it is *externally Bayesian*; i.e., finding the consensus distribution commutes with the process of revising distributions using a commonly agreed likelihood distribution. Next, we present consensus theorems using the LogOP.

Assumption 5. The local estimated pdf at the start of the consensus stage is positive everywhere, i.e., $\mathcal{F}^j_{k,0} = p^j_{k,0}(\boldsymbol{x}_k) > 0, \forall \boldsymbol{x}_k \in \mathcal{X}, \forall j \in \{1,\dots,m\}$. □

Assumption 5 is introduced to avoid regions with zero probability, since they would constitute vetoes and unduly great emphasis would get placed on them. Moreover, the LogOP guarantees that $\mathcal{F}^j_{k,\nu}$ will remain positive for all subsequent consensus loop.

Definition 1. *($\mathcal{H}^j_{k,\nu}$ vector for LogOP)* For the purpose of analysis, let us choose $\boldsymbol{x}_{k0} \in \mathcal{X}$ such that $p^j_{k,\nu}(\boldsymbol{x}_{k0}) > 0, \forall j \in \{1,\dots,m\}, \forall \nu \in \mathbb{N}$. Let us define $\mathcal{H}^j_{k,\nu} := \ln\left[\frac{p^j_{k,\nu}(\boldsymbol{x}_k)}{p_{k,\nu}(\boldsymbol{x}^j_{k0})}\right]$. Under Assumption 5, $\mathcal{H}^j_{k,\nu}$ is a well-defined function, but need not be a $\mathcal{L}_1$ function. Then, by simple algebraic manipulation of (7), we get:

$$\mathcal{H}^j_{k,\nu} = \sum_{\ell \in \mathcal{J}^j_k} a^{j\ell}_{k,\nu-1}\mathcal{H}^\ell_{k,\nu-1}, \forall j \in \{1,\dots,m\}, \nu \in \mathbb{N}, \quad (8)$$

which is similar to the LinOP (6). Let $\mathcal{U}_{k,\nu} := \left(\mathcal{H}^1_{k,\nu},\dots,\mathcal{H}^m_{k,\nu}\right)^T$ be an array of the estimates of all the agents during the ν^{th} consensus loop at the k^{th} time instant, then the equation (8) can be expressed concisely as:

$$\mathcal{U}_{k,\nu} = P_{k,\nu-1}\mathcal{U}_{k,\nu-1}, \forall \nu \in \mathbb{N}, \quad (9)$$

where $P_{k,\nu-1}$ is a matrix with entries $a^{jl}_{k,\nu-1}$. □

Thus we are able to use the highly nonlinear LogOP for combining the pdf estimates, but we have reduced the complexity of the problem to that of consensus using the LinOP. Next, we discuss the algorithm for achieving global exponential convergence on balanced graphs using the LogOP.

Assumption 6. The communication network topology of the multi–agent system $\mathcal{G}_k$ is strongly connected (SC) and balanced. The weighting factors $a^{j\ell}_{k,\nu-1}, \forall j,\ell \in \{1,\dots,m\}$ and the matrix $P_{k,\nu-1}$ have the following properties: (i) the weighting factors are the same for all consensus loops within each time instant, (ii) the matrix P_k conforms with the graph $\mathcal{G}_k$, (iii) the matrix P_k is row stochastic, and (iv) the weighting factors $a^{j\ell}_k$ are such that the digraph $\mathcal{G}_k$ is balanced. □

Theorem 2. (Consensus using the LogOP on SC Balanced Digraphs) *Under Assumption 5 and 6, using the LogOP (7), each $\mathcal{F}^j_{k,\nu}$ globally exponentially converges pointwise to the pdf $\mathcal{F}^\star_k$ given by:*

$$\mathcal{F}^\star_k = p^\star_k(\boldsymbol{x}_k) = \frac{\prod_{i=1}^m \left(p^i_{k,0}(\boldsymbol{x}_k)\right)^{\frac{1}{m}}}{\int_{\mathcal{X}} \prod_{i=1}^m \left(p^i_{k,0}(\boldsymbol{x}_k)\right)^{\frac{1}{m}} \, d\mu(\boldsymbol{x})} \quad (10)$$

at a rate faster or equal to $\sqrt{\lambda_{m-1}(P_k^T P_k)} = \sigma_{m-1}(P_k)$. Furthermore, their induced measures globally exponentially converge in total variation, i.e., $\lim_{\nu\to\infty} \mu_{\mathcal{F}^j_{k,\nu}} \xrightarrow{T.V.} \mu_{\mathcal{F}^\star_k}, \forall j \in \{1,\dots,m\}$.

Proof: See Appendix A. ■

The KL divergence is a measure of the information lost when the consensual pdf is used to approximate the locally estimated posterior pdfs. We now show that the consensual pdf $\mathcal{F}^\star_k$ obtained using Theorem 2, which is the weighted geometric average of the locally estimated posterior pdfs $\mathcal{F}^j_{k,0}, \forall j \in \{1,\dots,m\}$, minimizes the information lost during the consensus stage because it minimizes the sum of KL divergences with those pdfs.

Theorem 3. *The consensual pdf $\mathcal{F}^\star_k$ given by (10) globally minimizes the sum of Kullback–Leibler (KL) divergences with the locally estimated posterior pdfs at the start of the consensus stage $\mathcal{F}^j_{k,0}, \forall j \in \{1,\dots,m\}$, i.e.,*

$$\mathcal{F}^\star_k = \arg\min_{\rho \in \mathscr{L}_1(\mathcal{X})} \sum_{i=1}^m D_{KL}\left(\rho || \mathcal{F}^i_{k,0}\right), \quad (11)$$

where $\mathscr{L}_1(\mathcal{X})$ is the set of all pdfs over the state space $\mathcal{X}$ satisfying Assumption 5.

Proof: The sum of the KL divergences of a pdf $\rho \in \mathscr{L}_1(\mathcal{X})$ with the locally estimated posterior pdfs is given by:

$$\sum_{i=1}^m D_{KL}\left(\rho || \mathcal{F}^i_{k,0}\right) =$$
$$\sum_{i=1}^m \int_{\mathcal{X}} \left(\rho(\boldsymbol{x}_k)\ln(\rho(\boldsymbol{x}_k)) - \rho(\boldsymbol{x}_k)\ln(p^i_{k,0}(\boldsymbol{x}_k))\right) d\mu(\boldsymbol{x}). \quad (12)$$

Taking the logarithm, in the KL divergence formula, to the base $c := \left(\int_{\mathcal{X}} \prod_{i=1}^m \left(p^i_{k,0}(\boldsymbol{x}_k)\right)^{\frac{1}{m}} d\mu(\boldsymbol{x})\right)$ and then differentiating $\sum_{i=1}^m D_{KL}\left(\rho || \mathcal{F}^i_{k,0}\right)$ with respect to ρ using Leibniz integral rule [36, pp. 372] gives:

$$\sum_{i=1}^m \int_{\mathcal{X}} \left(\log_c(\rho(\boldsymbol{x}_k)) + 1 - \log_c(p^i_{k,0}(\boldsymbol{x}_k))\right) d\mu(\boldsymbol{x}) = 0\,,$$

which is minimized by $\mathcal{F}^\star_k$. ■

Note that if a central agent receives all the locally estimated posterior pdfs ($\mathcal{F}^j_{k,0}, \forall j \in \{1,\dots,m\}$) and is tasked to find the best estimate in the information theoretic sense, then it would also yield the same consensual pdf $\mathcal{F}^\star_k$ given by (10). Hence we claim to have achieved distributed estimation using this algorithm.

B. Communicating Probability Distributions

The consensus algorithms need the estimated pdfs to be communicated to other agents in the network. The first approach involves approximating the pdf by a weighted sum of Gaussians and then transmitting this approximate distribution. Let $\mathcal{N}(\boldsymbol{x}_k - m_i, B_i)$ denote the Gaussian density function, where the mean is the n_x-vector m_i and the covariance is the positive-definite symmetric matrix B_i. The Gaussian sum approximations lemma of [38, pp. 213] states that any pdf $\mathcal{F} = p(\boldsymbol{x}_k)$ can be approximated as closely as desired in the $\mathscr{L}_1(\mathbb{R}^{n_x})$ space by a pdf of the form $\hat{\mathcal{F}} = \hat{p}(\boldsymbol{x}_k) = \sum_{i=1}^{n_g} \alpha_i \mathcal{N}(\boldsymbol{x}_k - m_i, B_i)$, for some integer n_g and positive scalars α_i with $\sum_{i=1}^{n_g} \alpha_i = 1$. For an acceptable communication error $\varepsilon_{\text{comm}} > 0$, there exists n_g, α_i, m_i and B_i such that $\|\mathcal{F} - \hat{\mathcal{F}}\|_{\mathcal{L}_1} \leq \varepsilon_{\text{comm}}$. Let $\tilde{\mathcal{F}}^j_{k,\nu}$ be the LinOP solution after combining local pdfs corrupted by communication error, i.e., $\tilde{\mathcal{F}}^j_{k,\nu} := \mathcal{T}\left(\cup_{\ell \in \mathcal{J}^j_k} \{\hat{\mathcal{F}}^\ell_{k,\nu-1}\}\right)$ where $\mathcal{T}(\cdot)$ is either LinOP or LogOP. Then we get $\|\mathcal{F}^j_{k,\nu} - \tilde{\mathcal{F}}^j_{k,\nu}\|_{\mathcal{L}_1} \leq \nu\varepsilon_{\text{comm}}, \forall \nu \in \mathbb{N}$, where $\mathcal{F}^j_{k,\nu}$ is the true solution obtained from uncorrupted local pdfs. Several techniques for estimating the parameters are discussed in the Gaussian mixture model literature, like maximum likelihood (ML) and maximum a posteriori (MAP) parameter estimation [39]– [41]. Hence, in order to communicate the pdf $\hat{\mathcal{F}}$, the agent needs to transmit $\frac{1}{2} n_g n_x (n_x + 3)$ real numbers.

If particle filters are used to evaluate the Bayesian filter and combine the pdfs [22], then the resampled particles represent the agent's estimated pdf of the target. Hence communicating pdfs is equivalent to transmitting these resampled particles. The information theoretic approach for communicating pdfs is discussed in [42]. Now that we have established that communication of pdfs is possible, let us discuss the complete BCF algorithm.

Algorithm 1 BCF–LogOP on SC Balanced Digraphs

(one cycle of j^{th} agent during k^{th} time instant)
Given the pdf from previous time step
  $\mathcal{F}^j_{k-1} = p^j_{k-1}(\boldsymbol{x}_{k-1})$
Set n_{loop}, the weights $a^{j\ell}_k$ } Theorems 2, 4
while tracking **do**
  Compute the prior pdf $p^j_k(\boldsymbol{x}_k)$
  Compute the posterior pdf $\mathcal{F}^j_k$
  } Bayesian Filtering Stage
  for $\nu = 1$ to n_{loop}
    if $\nu = 1$ **then** Set $\mathcal{F}^j_{k,0} = \mathcal{F}^j_k$
    end if
    Obtain the communicated pdfs $\mathcal{F}^\ell_{k,\nu-1}, \forall \ell \in \mathcal{J}^j_k$
    Compute the new pdf $\mathcal{F}^j_{k,\nu}$
  end for
  } LogOP–based Consensus Stage
  Set $\mathcal{F}^j_k = \mathcal{F}^j_{k,n_{\text{loop}}}$ **end while**

IV. Main Algorithms: Bayesian Consensus Filtering

In this section, we finally present the BCF algorithm illustrated by **Algorithm 1**. The BCF is performed in two steps: (i) each agent locally estimates the pdf of the target's states using a Bayesian filter with/without measurements from neighboring agents, as discussed in Section II-A, and (ii) during the consensus stage, each agent recursively transmits its pdf estimate of the target's states to other agents, receives estimates of its neighboring agents, and combines them using the LogOP as discussed in Section III-A. In this section, we compute the number of consensus loops (n_{loop}) needed to reach a satisfactory consensus estimate across the network and discuss the convergence of this algorithm.

Definition 2. (*Disagreement vector* $\boldsymbol{\theta}_{k,\nu}$) Let us define $\boldsymbol{\theta}_{k,\nu} := \left(\theta^1_{k,\nu}, \ldots, \theta^m_{k,\nu}\right)^T$, where $\theta^j_{k,\nu} := \|\mathcal{F}^j_{k,\nu} - \mathcal{F}^\star_k\|_{\mathcal{L}_1}$. Since the $\mathcal{L}_1$ distance between pdfs is upper bounded by 2, the ℓ_2 norm of the disagreement vector ($\|\boldsymbol{\theta}_{k,\nu}\|_{\ell_2}$) is upper bounded by $2\sqrt{m}$. □

This conservative bound is used to obtain the minimum number of consensus loops for achieving ε-consensus across the network, while tracking a moving target. Let us now quantify the divergence of the local pdfs during the Bayesian filtering stage.

Definition 3. (*Error propagation dynamics* $\boldsymbol{\Gamma}(\cdot)$) Let us assume that the dynamics of the ℓ_2 norm of the disagreement vector during the Bayesian filtering stage can be obtained from the target dynamics and measurement models (1) and (2). The error propagation dynamics $\boldsymbol{\Gamma}(\cdot)$ estimates the maximum divergence of the local pdfs during the Bayesian filtering stage, i.e., $\|\boldsymbol{\theta}_{k,0}\|_{\ell_2} \leq \boldsymbol{\Gamma}\left(\|\boldsymbol{\theta}_{k-1,n_{\text{loop}}}\|_{\ell_2}\right)$, where $\|\boldsymbol{\theta}_{k-1,n_{\text{loop}}}\|_{\ell_2}$ is the disagreement vector with respect to $\mathcal{F}^\star_{k-1}$ at the end of the consensus stage during the $(k-1)^{\text{th}}$ time instant; and $\|\boldsymbol{\theta}_{k,0}\|_{\ell_2}$ is the disagreement vector with respect to $\mathcal{F}^\star_k$ after the update stage during the k^{th} time instant. □

Next we obtain the minimum number of consensus loops for achieving ε-consensus across the network and also derive conditions on the communication network topology for a given number of consensus loops.

Theorem 4. (BCF–LogOP on SC Balanced Digraphs) *Under Assumptions 5, 6, and in the absence of communication inaccuracies, each agent tracks the target using the BCF algorithm. For some acceptable consensus error* $\varepsilon_{\text{consensus}} > 0$ *and* $\gamma_k = \min\left(\boldsymbol{\Gamma}\left(\|\boldsymbol{\theta}_{k-1,n_{\text{loop}}}\|_{\ell_2}\right), 2\sqrt{m}\right)$*:*
(i) if the number of consensus loops is at least $n_{\text{loop}} \geq \left\lceil \frac{\ln(\varepsilon_{\text{consensus}}/\gamma_k)}{\ln \sigma_{m-1}(P_k)} \right\rceil$ *for a given* P_k*; or*
(ii) if the communication network topology (P_k) *during the* k^{th} *time instant is such that* $\sigma_{m-1}(P_k) \leq \left(\frac{\varepsilon_{\text{consensus}}}{\gamma_k}\right)^{\frac{1}{n_{\text{loop}}}}$ *for a given* n_{loop}*;*
then the ℓ_2 *norm of the disagreement vector at the end of the consensus stage is less than* $\varepsilon_{\text{consensus}}$*, i.e.,* $\|\boldsymbol{\theta}_{k,n_{\text{loop}}}\|_{\ell_2} \leq \varepsilon_{\text{consensus}}$.

Proof: It follows from Theorem 2 that, if $\boldsymbol{\theta}_{k,0}$ is the initial disagreement vector at the start of the consensus stage, then $\|\boldsymbol{\theta}_{k,n_{\text{loop}}}\|_{\ell_2} \leq (\sigma_{m-1}(P_k))^{n_{\text{loop}}} \|\boldsymbol{\theta}_{k,0}\|_{\ell_2} \leq (\sigma_{m-1}(P_k))^{n_{\text{loop}}} \gamma_k$. Thus, we get the conditions on n_{loop}

or $\sigma_{m-1}(P_k)$ from the inequality $(\sigma_{m-1}(P_k))^{n_{\text{loop}}}\gamma_k \leq \varepsilon_{\text{consensus}}$. ■

Note that for the particular case where $n_{\text{loop}}=1$, we need $\sigma_{m-1}(P_k) \leq \frac{\varepsilon_{\text{consensus}}}{\gamma_k}$ for ε-convergence across the network.

A. Hierarchical Bayesian Consensus Filtering

In this section, we modify the original problem statement such that only m_1 out of m agents are able to observe the target at the k^{th} time instant. In this scenario, the other $m_2(=m-m_1)$ agents are not able to observe the target. Without loss of generality, we assume that the first m_1 agents, i.e., $\{1,2,\ldots,m_1\}$, are tracking the target. During the Bayesian filtering stage, each *tracking agent* (i.e., agent tracking the target) estimates the posterior pdf of the target's states at the k^{th} time instant ($\mathcal{F}_k^j = p_k^j(\boldsymbol{x}_k|\boldsymbol{z}_k^{\mathcal{S}_k^j\cap\{1,\ldots m_1\}}), \forall j \in \{1,\ldots,m_1\}$) using the estimated prior pdf of the target's states ($\mathcal{F}_{k-1}^j$) and the new measurement array $\boldsymbol{z}_k^{\mathcal{S}_k^j\cap\{1,\ldots m_1\}} := \left\{\boldsymbol{z}_k^\ell, \forall \ell \in \mathcal{S}_k^j \cap \{1,\ldots m_1\}\right\}$ obtained from the neighboring tracking agents. Each *non-tracking agent* (i.e., agent not tracking the target) only propagates its prior pdf during this stage to obtain $p_k^j(\boldsymbol{x}_k), \forall j \in \{m_1+1,\ldots,m\}$.

The objective of hierarchical consensus algorithm is to guarantee pointwise convergence of each $\mathcal{F}_{k,\nu}^j, \forall j \in \{1,\ldots,m\}$ to a pdf $\mathcal{F}_k^\star$ and only the local estimates of the agents tracking the target contribute to the consensual pdf. This is achieved by each tracking agent recursively transmitting its estimate of the target's states to other agents, only receiving estimates from its neighboring tracking agents and updating its estimate of the target. On the other hand, each non-tracking agent recursively transmits its estimate of the target's states to other agents, receives estimates from all its neighboring agents and updates its estimate of the target. Let $\mathcal{D}_k$ represent the communication network topology of only the tracking agents.

Assumption 7. The communication network topologies $\mathcal{G}_k$ and $\mathcal{D}_k$ are SC and $\mathcal{D}_k$ is balanced. The weighting factors $a_{k,\nu-1}^{j\ell}, \forall j,\ell \in \{1,\ldots,m\}$ and the matrix $P_{k,\nu-1}$ have the following properties: (i) the weighting factors are the same for all consensus loops within each time instants; (ii) if $j \in \{1,\ldots,m_1\}$, then $a_k^{j\ell}>0$ if and only if $\ell \in \mathcal{J}_k^j \cap \{1,\ldots,m_1\}$, else $a_k^{j\ell}=0$; (iii) if $j \in \{m_1+1,\ldots,m\}$, then $a_k^{j\ell}>0$ if and only if $\ell \in \mathcal{J}_k^j$, else $a_k^{j\ell}=0$; (iv) the matrix P_k is row stochastic, and (v) the weighting factors $a_k^{j\ell}$ are such that the digraph $\mathcal{D}_k$ is balanced. □

Theorem 5. (Hierarchical Consensus using the LogOP on SC Balanced Digraphs) *Under Assumptions 5 and 7, using the LogOP (7), each $\mathcal{F}_{k,\nu}^j$ globally exponentially converges pointwise to the pdf $\mathcal{F}_k^\star$ given by:*

$$\mathcal{F}_k^\star = p_k^\star(\boldsymbol{x}_k) = \frac{\prod_{i=1}^{m_1}\left(p_{k,0}^i(\boldsymbol{x}_k)\right)^{\frac{1}{m_1}}}{\int_{\mathcal{X}}\prod_{i=1}^{m_1}\left(p_{k,0}^i(\boldsymbol{x}_k)\right)^{\frac{1}{m_1}}\, d\mu(\boldsymbol{x})} \tag{13}$$

at a rate faster or equal to $\sqrt{\lambda_{m_1-1}(P_{k1}^T P_{k1})} = \sigma_{m_1-1}(P_{k1})$, where the matrix P_{k1} conforms with the directed graph $\mathcal{D}_k$. Only the initial estimates of the tracking agents contribute to the consensual pdf $\mathcal{F}_k^\star$. Furthermore, their induced measures converge in total variation, i.e., $\lim_{\nu\to\infty}\mu_{\mathcal{F}_{k,\nu}^j} \xrightarrow{T.V.} \mu_{\mathcal{F}_k^\star}, \forall j \in \{1,\ldots,m\}$.

Proof: The proof follows from Theorem 2. In fact, the inessential states die out geometrically fast [43, Theorem 4.3, pp. 120]. ■

A simulation example of Hierarchical BCF–LogOP algorithm for tracking orbital debris in space is discussed in the next section.

V. Numerical Example

Currently, there are over ten thousand objects in Earth orbit, of size 0.5 cm or greater, and almost 95% of them are nonfunctional space debris. These debris pose a significant threat to functional spacecraft and satellites in orbit. The US has established the Space Surveillance Network (SSN) for ground based observations of the orbital debris using radars and optical telescopes [44] (See Fig. 1). In February 2009, the Iridium–33 satellite collided with the Kosmos–2251 satellite and a large number of debris fragments were created. In this section, we use the Hierarchical BCF–LogOP Algorithm to track one of the Iridium–33 debris created in this collision.

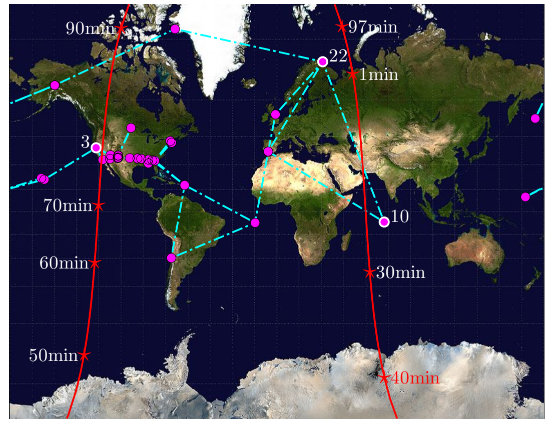

Fig. 1. The SSN locations are shown along with their static SC balanced communication network topology. The orbit of the Iridium–33 debris is shown in red, where ★ marks its actual position during particular time instants.

The actual two-line element set (TLE) of the Iridium–33 debris was accessed from North American Aerospace Defense Command (NORAD) on 4th Dec 2013. The nonlinear Simplified General Perturbations (SGP4) model, which uses an extensive gravitational model and accounts for the drag effect on mean motion [45], [46], is used as the target dynamics model. If the debris is visible above the sensor's horizon, then it is assumed to create a single measurement during each time step of one minute. The heterogeneous measurement model of the j^{th} sensor is given by:

$$\boldsymbol{z}_k^j = \boldsymbol{x}_k + \boldsymbol{w}_k^j, \text{ where } \boldsymbol{w}_k^j = \mathcal{N}\left(0, (1000+50j)\times\mathbf{I}\right),$$

where $\boldsymbol{x}_k \in \mathbb{R}^3$ is the actual location of the debris. Since it is not possible to implement the SGP4 target dynamics on

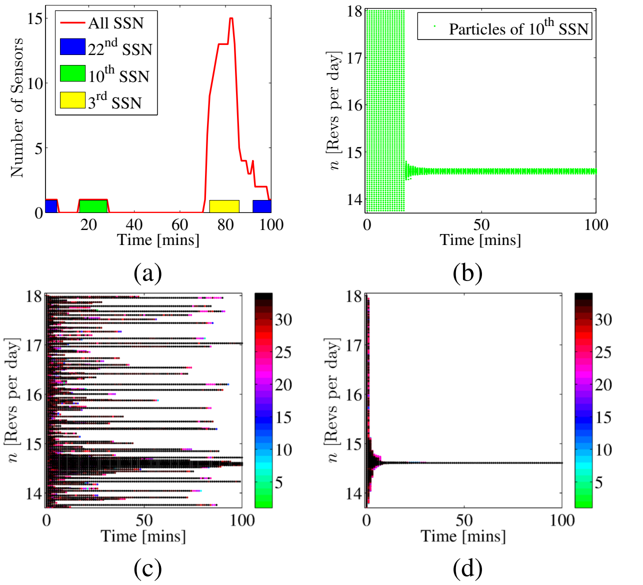

Fig. 2. (a) Number of SSN sensors observing debris. (b) Trajectories of particles for stand-alone Bayesian filters for 10th SSN sensor. Trajectories of particles of all sensors for (c) Hierarchical BCF–LinOP and (d) Hierarchical BCF–LogOP.

distributed estimation algorithms discussed in the literature [1]– [15], we compare the performance of our Hierarchical BCF–LogOP algorithm against the Hierarchical BCF–LinOP algorithm, where the LinOP is used during the consensus stage.

In this simulation example, we simplify the debris tracking problem by assuming only the mean motion (n) of the debris is unknown, which needs to be estimated within 100 minutes. Hence, each sensor knows the other TLE parameters of the debris and an uniform prior distribution ($\mathcal{F}_0^j$) is assumed. Note that at any time instant, only a few of the SSN sensors can observe the debris, as shown in Fig 2(a). The estimates of the stand-alone Bayesian filter for the 10th sensor do not converge due to large measurement errors, in spite of observing the debris for some time.

Particle filters with resampling are used to evaluate the Bayesian filters and communicate pdfs in the Hierarchical BCF algorithms. 100 particles are used by each sensor and 10 consensus loops are executed during each time step of one minute. As expected, all the sensors converge on the correct value of n of 14.6 revs per day. The Hierarchical BCF–LinOP estimates are multimodal for the first 90 minutes. On the other hand, the Hierarchical BCF–LogOP estimates converges to the correct value within the first 10 minutes because the LogOP algorithm efficiently communicates the best consensual estimate to other sensors during each time step and achieves consensus across the network.

VI. Conclusion

In this paper, we extended the scope of distributed estimation algorithms in a Bayesian filtering framework in order to simultaneously track targets, with general nonlinear time-varying target dynamic models, using a strongly connected network of heterogeneous agents, with general nonlinear time-varying measurement models. The LogOP algorithm on SC balanced digraph converges globally exponentially, and the consensual pdf minimizes the information lost during the consensus stage because it minimizes the sum of KL divergences to each locally estimated probability distribution. We introduced the BCF algorithm, where the local estimated posterior pdfs of the target's states are first updated using the Bayesian filter and then recursively combined during the consensus stage using LogOP, so that the agents can track a moving target and also maintain consensus across the network. Conditions for exponential convergence of the BCF algorithm and constraints on the communication network topology have been studied. The Hierarchical BCF algorithm, where some of the agents do not observe the target, has also been investigated. Simulation results demonstrate the effectiveness of the BCF algorithms for nonlinear distributed estimation problems.

Appendix A
Proof of Theorem 2

Under Assumption 6, P_k is a nonnegative, doubly stochastic and irreducible matrix. Hence Perron–Frobenius theorem (cf. [43, pp. 3]) states that $\lim_{\nu\to\infty} P_k^{\nu} = \frac{1}{m}\mathbf{1}\mathbf{1}^T$ and each $\mathcal{H}_{k,\nu}^j$ converges pointwise to $\mathcal{H}_k^\star = \frac{1}{m}\mathbf{1}^T\mathcal{U}_{k,0} = \frac{1}{m}\sum_{i=1}^m \mathcal{H}_{k,0}^i$. We have $\forall \boldsymbol{x}_k \in \mathcal{X}$:

$$\lim_{\nu\to\infty}\left(\ln p_{k,\nu}^j(\boldsymbol{x}_k) - \ln p_{k,\nu}^j(\boldsymbol{x}_{k0})\right) = \ln p_k^\star(\boldsymbol{x}_k) - \ln p_k^\star(\boldsymbol{x}_{k0}).$$

As $\exists \bar{\boldsymbol{x}}_{k0} \in \mathcal{X}$ such that $\lim_{\nu\to\infty} p_{k,\nu}^j(\bar{\boldsymbol{x}}_{k0}) = p_k^\star(\bar{\boldsymbol{x}}_{k0})$, we get $\lim_{\nu\to\infty} p_{k,\nu}^j(\boldsymbol{x}_k) = p_k^\star(\boldsymbol{x}_k), \forall \boldsymbol{x}_k \in \mathcal{X}$. Thus each $\mathcal{F}_{k,\nu}^j$ converges pointwise to the consensual pdf $\mathcal{F}_k^\star$ given by (10). By Lemma 1, the measure induced by $\mathcal{F}_{k,\nu}^j$ on $\mathscr{X}$ converges in total variation to the measure induced by $\mathcal{F}_k^\star$ on $\mathscr{X}$, i.e., $\lim_{\nu\to\infty} \mu_{\mathcal{F}_{k,\nu}^j} \xrightarrow{\text{T.V.}} \mu_{\mathcal{F}_k^\star}$.

If $V_{\text{tr}} = \left[\frac{1}{\sqrt{m}}\mathbf{1}, V_{\text{s}}\right]$ are the orthonormal eigenvectors of $P_k^T P_k$, then by spectral decomposition [47] we get that the rate at which $\mathcal{U}_{k,\nu}$ synchronizes to $\frac{1}{\sqrt{m}}\mathbf{1}$ (or $\mathcal{U}_k^\star$) is equal to the rate at which $V_{\text{s}}^T\mathcal{U}_{k,\nu} \to \mathbf{0}^{(m-1)\times 1}$. If $\Phi_{k,\nu} = (V_{\text{s}}^T\mathcal{U}_{k,\nu})^T V_{\text{s}}^T\mathcal{U}_{k,\nu}$ is a candidate Lyapunov function, then $\Phi_{k,\nu} \leq \left(\lambda_{\max}(V_{\text{s}}^T P_k^T P_k V_s)\right)\Phi_{k,\nu-1}$. Hence each $\mathcal{H}_{k,\nu}^j$ globally exponentially converges pointwise to $\mathcal{H}_k^\star$ with a rate faster or equal to $\sqrt{\lambda_{m-1}(P_k^T P_k)} = \sigma_{m-1}(P_k)$.

Next, we need to find the rate of convergence of $\mathcal{F}_{k,\nu}^j$ to $\mathcal{F}_k^\star$. Let us define the continuous function $\alpha_{k,\nu}^j(\boldsymbol{x}_k)$ such that $\alpha_{k,\nu}^j(\boldsymbol{x}_k) = \left[\frac{p_{k,\nu}^j(\boldsymbol{x}_k)}{p_k^\star(\boldsymbol{x}_k)}\frac{p_k^\star(\boldsymbol{x}_{k0})}{p_{k,\nu}^j(\boldsymbol{x}_{k0})}\right]$ if $p_{k,\nu}^j(\boldsymbol{x}_k)p_k^\star(\boldsymbol{x}_{k0}) \geq p_k^\star(\boldsymbol{x}_k)p_{k,\nu}^j(\boldsymbol{x}_{k0})$ and $\alpha_{k,\nu}^j(\boldsymbol{x}_k) = \left[\frac{p_k^\star(\boldsymbol{x}_k)}{p_{k,\nu}^j(\boldsymbol{x}_k)}\frac{p_{k,\nu}^j(\boldsymbol{x}_{k0})}{p_k^\star(\boldsymbol{x}_{k0})}\right]$ otherwise. Then we get:

$$\alpha_{k,\nu}^j(\boldsymbol{x}_k) \leq \left(\alpha_{k,0}^j(\boldsymbol{x}_k)\right)^{(\sigma_{m-1}(P_k))^\nu}. \tag{14}$$

Using the mean value theorem, (14) can be simplified to:

$$\alpha_{k,\nu}^j(\boldsymbol{x}_k) - 1 \leq (\sigma_{m-1}(P_k))^\nu\left(\alpha_{k,0}^j(\boldsymbol{x}_k) - 1\right). \tag{15}$$

Irrespective of the orientation of $\alpha^j_{k,\nu}(\boldsymbol{x}_k)$ and $\alpha^j_{k,0}(\boldsymbol{x}_k)$, (15) can be written as (16) by multiplying with $\frac{1}{\alpha^j_{k,\nu}(\boldsymbol{x}_k)}$ or $\frac{1}{\alpha^j_{k,0}(\boldsymbol{x}_k)}$, and then with $p^\star_k(\boldsymbol{x}_k)$.

$$\begin{aligned}&\left|\frac{p^\star_k(\boldsymbol{x}_{k0})}{p^j_{k,\nu}(\boldsymbol{x}_{k0})}p^j_{k,\nu}(\boldsymbol{x}_k) - p^\star_k(\boldsymbol{x}_k)\right| \\ &\leq (\sigma_{m-1}(P_k))^\nu \left|\frac{p^\star_k(\boldsymbol{x}_{k0})}{p^j_{k,0}(\boldsymbol{x}_{k0})}p^j_{k,0}(\boldsymbol{x}_k) - p^\star_k(\boldsymbol{x}_k)\right| . \end{aligned} \tag{16}$$

If we choose $\tilde{\boldsymbol{x}}_{k0} \in \mathcal{X}$ such that $p^j_{k,0}(\tilde{\boldsymbol{x}}_{k0}) = p^\star_k(\tilde{\boldsymbol{x}}_{k0})$, then we can simplify (16) to:

$$\left|p^j_{k,\nu}(\boldsymbol{x}_k) - p^\star_k(\boldsymbol{x}_k)\right| \leq (\sigma_{m-1}(P_k))^\nu \left|p^j_{k,0}(\boldsymbol{x}_k) - p^\star_k(\boldsymbol{x}_k)\right| .$$

Thus each $\mathcal{F}^j_{k,\nu} = p^j_{k,\nu}(\boldsymbol{x}_k)$ globally exponentially converges to $\mathcal{F}^\star_k = p^\star_k(\boldsymbol{x}_k)$ with a rate faster or equal to $\sigma_{m-1}(P_k)$. ■

References

[1] V. Borkar and P. Varaiya, "Asymptotic agreement in distributed estimation," *IEEE Trans. Autom. Control*, vol. 27, no. 3, pp. 650 – 655, 1982.

[2] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Sel. Topics Signal Process.*, vol. 5, pp. 674 – 690, Aug 2011.

[3] I. Schizas, G. Mateos, and G. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. Signal Process.*, vol. 57, pp. 2365 – 2382, June 2009.

[4] D. Spanos, R. Olfati-Saber, and R. M. Murray, "Distributed sensor fusion using dynamic consensus," in *Proc. IFAC*, 2005.

[5] M. Coates, "Distributed particle filters for sensor networks," in *Proc. 3rd Int. Symp. Inform. Process. Sensor Networks*, (New York, USA), pp. 99–107, 2004.

[6] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520 – 1533, 2004.

[7] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988 – 1001, 2003.

[8] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, pp. 655 – 661, May 2005.

[9] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, pp. 65 – 78, 2004.

[10] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803 – 812, 1986.

[11] F. Zhang and N. E. Leonard, "Cooperative filters and control for cooperative exploration," *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 650–663, 2010.

[12] N. Ahmed, J. Schoenberg, and M. Campbell, *Robotics: Science and Systems VIII*, ch. Fast Weighted Exponential Product Rules for Robust General Multi-Robot Data Fusion, pp. 9–16. MIT Press, 2013.

[13] M. Demetriou and D. Uciński, "State estimation of spatially distributed processes using mobile sensing agents," in *Amer. Control Conf.*, (San Francisco, CA, USA), pp. 1770–1776, 2011.

[14] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. of the IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.

[15] B. Açikmeşe, F. Y. Hadaegh, D. P. Scharf, and S. R. Ploen, "Formulation and analysis of stability for spacecraft formations," *IET Control Theory Appl.*, vol. 1, no. 2, pp. 461–474, 2007.

[16] Y. Xu, V. Gupta, and C. Fischione, "Distributed estimation," in *E-reference Signal Processing*, Elsevier, 2013. Editors: Rama Chellappa, Sergios Theodoridis.

[17] R. Olfati-Saber, "Kalman-consensus filter : Optimality, stability, and performance," in *48th IEEE Conf. Decision Control*, (Shanghai, China), pp. 7036–7042, December 2009.

[18] M. H. DeGroot, *Probability and Statistics*. Cambridge, Massachusetts: Addison-Wesley, 1975.

[19] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[20] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, Massachusetts: The MIT Press, 2005.

[21] C. Boutilier, T. Dean, and S. Hanks, "Decision-theoretic planning: Structural assumptions and computational leverage," *J. Artificial Intell. Research*, vol. 11, pp. 1–94, 1999.

[22] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, pp. 174–188, February 2002.

[23] O. Lebeltel, P. Bessiere, J. Diard, and E. Mazer, "Bayesian robot programming," *Autonomous Robots*, vol. 16, pp. 49–79, January 2004.

[24] M. H. DeGroot, "Reaching a consensus," *J. Amer. Statistical Assoc.*, vol. 69, no. 345, pp. 688 – 704, 1960.

[25] C. Genest and J. V. Zidek, "Combining probability distributions: A critique and an annotated bibliography," *Statistical Sci.*, vol. 1, no. 1, pp. 114 – 135, 1986.

[26] S. Chatterjee and E. Seneta, "Towards consensus: Some convergence theorems on repeated averaging," *J. Appl. Probability*, vol. 14, no. 1, pp. 89 – 97, 1977.

[27] P. Velagapudi, O. Prokopyev, K. Sycara, and P. Scerri, "Maintaining shared belief in a large multiagent team," in *Proc. of FUSION*, 2007.

[28] S. Yüksel, "Stochastic nestedness and the belief sharing information pattern," *IEEE Trans. Autom. Control*, vol. 54, no. 12, pp. 2773–2786, 2009.

[29] C. S. R. Fraser, L. F. Bertuccelli, H.-L. Choi, and J. P. How, "A hyperparameter consensus method for agreement under uncertainty," *Automatica*, vol. 48, no. 2, pp. 374 – 380, 2012.

[30] O. Hlinka, O. Slučiak, F. Hlawatsch, P. M. Djuric, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4334–4349, 2012.

[31] G. Pavlin, P. Oude, M. Maris, J. Nunnink, and T. Hood, "A multi-agent systems approach to distributed Bayesian information fusion," *Inform. Fusion*, vol. 11, pp. 267–282, 2010.

[32] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, pp. 210–225, 2012.

[33] B. J. Julian, M. Angermann, M. Schwager, and D. Rus, "Distributed robotic sensor networks: An information-theoretic approach," *Inter. J. of Robotics Research*, vol. 31, no. 10, pp. 1134–1154, 2012.

[34] O. Punska, "Bayesian approaches to multi-sensor data fusion," Master's thesis, Dept. of Eng., Univ. of Cambridge, 1999.

[35] T. Zhao and A. Nehorai, "Distributed sequential Bayesian estimation of a diffusive source in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1511 – 1524, 2007.

[36] R. Durrett, *Probability: Theory and Examples*. Thomson Brooks, 2005.

[37] M. Bacharach, "Normal Bayesian dialogues," *J. Amer. Statistical Assoc.*, vol. 74, no. 368, pp. 837 – 846, 1979.

[38] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Mineola, New York: Dover Publications, 2005.

[39] D. A. Reynolds, "Gaussian mixture models," *Encyclopedia of Biometric Recognition*, February 2008.

[40] G. J. McLachlan and K. E. Basford, *Mixture models : inference and applications to clustering*. New York, N.Y.: M. Dekker, 1988.

[41] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Trans. Signal Process.*, vol. 51, pp. 2602–2612, Oct. 2003.

[42] G. Kramer and S. A. Savari, "Communicating probability distributions," *IEEE Trans. Inf. Theory*, vol. 53, pp. 518–525, February 2007.

[43] E. Seneta, *Non-negative Matrices and Markov Chains*. New York, NY: Springer-Verlag, 2006.

[44] D. A. Vallado and J. D. Griesbach, "Simulating space surveillance networks," in *AAS/AIAA Astrodynamics Specialist Conf.*, (Girdwood), 2012. Paper AAS 11-580.

[45] F. R. Hoots and R. L. Roehrich, "Spacetrack report number 3: Models for propagation of NORAD element sets," tech. rep., U.S. Air Force Aerospace Defense Command, Colorado Springs, CO., 1980.

[46] D. Vallado and P. Crawford, "SGP4 orbit determination," in *AIAA/AAS Astrodynamics Specialist Conf.*, 2008.

[47] S.-J. Chung, S. Bandyopadhyay, I. Chang, and F. Y. Hadaegh, "Phase synchronization control of complex networks of Lagrangian systems on adaptive digraphs," *Automatica*, vol. 49, pp. 1148–1161, May 2013.